Slide 1

Gelato@UNSW VM Work

Ian Wienand, Paul Davies, Lucy Chubb

This work supported by UNSW and HP through the Gelato Federation

Slide 2

About the Talk

- ➜ Work in progress talk
- ➜ Please talk to us!

Topics:

- ➜ Superpages
- ➜ Long Format VHPT
- ➜ Page Table Abstraction

Slide 3

Superpages

What?:

- ➜ Use page size > base page size
- ➜ `HugeTLB`
  - ✗ Programmer Visible
  - ✗ Confusing
  - ☑ Suits legacy architectures (x86/64, Power)

Why?:

- ➜ More coverage from TLB
- ➜ Transparency is a worthwhile, unsolved problem
- ➜ Itanium provides wide range of interesting MMU options
  - ➜ **With LVHPT!**

Slide 4

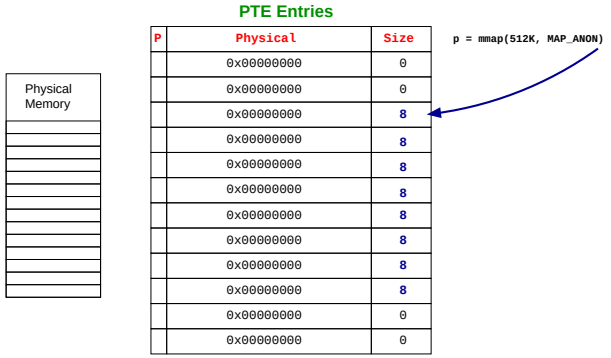
Our Approach

Cluster individual pages:

1. Tag entries with size when created - `mmap(MAP_ANONYMOUS)`
2. On fault, check page size flag
3. Backtrack and allocate physical superpage
4. Point each sub-page of superpage to actual physical page

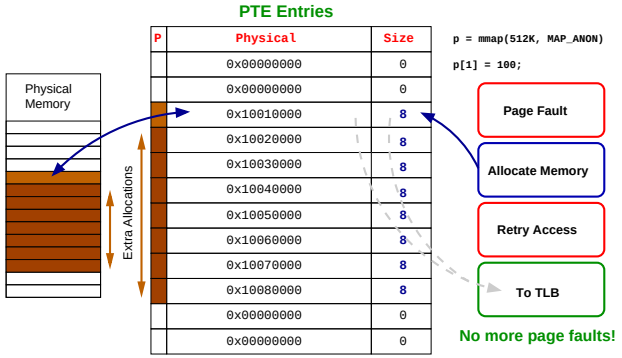
Slide 5

Superpage Overview



Slide 6

Superpage Overview



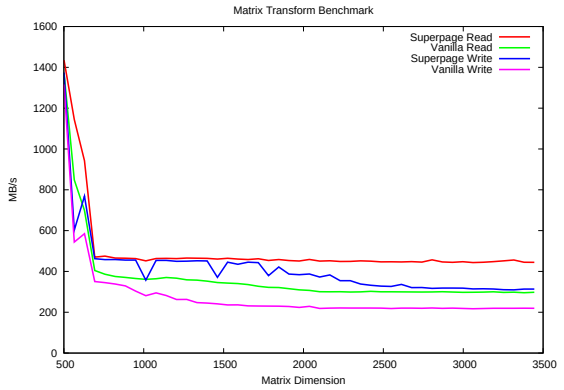
Slide 7

Issues

- Work with Itanium features
- Page Tables
  - Double Size PTE
  - Mapping across PMD's
  - Better approaches
- Fragmentation
  - External Fragmentation (Ozlabs, Rice)
  - Internal Fragmentation (Rice)
- Promotion/Demotion
  - Competitive Algorithms

Slide 8

Matrix Transformation



Slide 9

Dodgy Results

- `lmbench` - no massive differences
- `SPECint` - `mcf` shows good improvement
- Lee's benchmark?

Slide 10

Virtually Hashed Page Table

Problem:

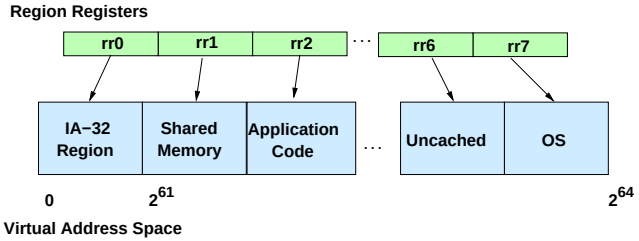
- TLB is fast but small
- Walking page tables via software is slow

Solution:

- Keep a "cache" of translations in RAM
- Let the hardware do the walking!
- The **VHPT**

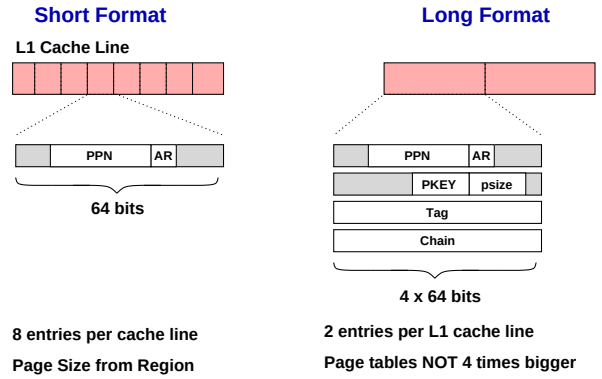
Slide 11

Fast VM Overview - Regions



Slide 12

Two types of VHPT



8 entries per cache line

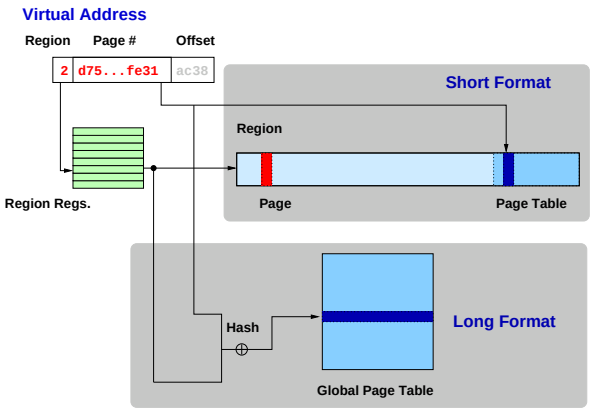
Page Size from Region

2 entries per L1 cache line

Page tables NOT 4 times bigger

Slide 13

VHPT Differences



Slide 14

Two types of VHPT

- ➜ **Short** : Cache Friendly, TLB Unfriendly
  - ➜ More entries in a cache line
  - ➜ Extra TLB entry per page of PTEs
- ➜ **Long** : Cache Unfriendly, TLB Friendly
  - ➜ Less entries per cache line
  - ➜ Global page table pinned with single entry

Why long?:

- ☑ Entries can hold page size, protection keys
- ☒ Entries are twice as big.
- ➜ Would like to talk to hackers and users.

Slide 15

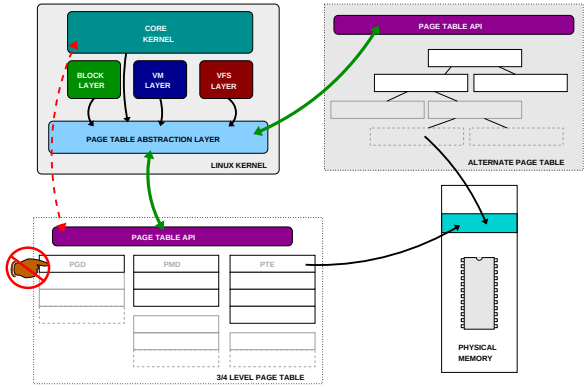
Page Table Abstraction

Problems with current implementations:

- ➜ 3 (4? 5? X?) level page table sub-optimal
- ➜ Main architectures have differing needs

Slide 16

Solution? Page Table Abstraction



Slide 17

Issues:

- ☑ Performance - not bad!
- ☑ Hardware walked page tables - we have ideas
- ☑ Alternate Implementations - GPT
- ☒ Convincing Linux community to rip apart VM - hard

Slide 18

WHERE DO YOU GET IT?

`http://www.gelato.unsw.edu.au/IA64Wiki`

or

for "`ia64 wiki`"

QUESTIONS?